OIPE JAN 2 2 2002 PATENT & TRADEMARK OFFICE

#6 2621

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Docket No. **9230.00**

Application of

Nicholas Heard

Serial No. **09/865,828**

Filed: **May 24, 2001**

RECEIVED

FEB 0 4 2002

Technology Center 2600

CLAIM FOR BENEFIT OF EARLIER-FILED FOREIGN APPLICATION

Confirmation No.: **5577**

Group Art Unit: **2621**

Examiner: **Unknown**

FOR: **METHOD AND APPARATUS FOR DETERMINING ONE OR MORE STATISTICAL ESTIMATORS OF CUSTOMER BEHAVIOR**

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231 on NOV 07 2001 (Date of Deposit).

Shirley Doll

Assistant Commissioner for Patents

Washington, D.C. 20231

Sir:

Applicants wish to claim the benefit of the filing date of the earlier G.B. Application Serial No. **0013011.2**, filed on **May 26, 2000**, recited in the Declaration under the provision of 35 U.S.C. 119, and accordingly, Applicants submit herewith a certified copy of said application.

Respectfully submitted,

Michael Chan
Reg. No. 33,663
Attorney for Applicant(s)

NCR Corporation, Law Department, WHQ5E
1700 S. Patterson Blvd., Dayton, OH 45479-0001
Tel. No. 937-445-4956/Fax No. 937-445-3733

NOV 07 2001

THIS PAGE BLANK (USPTO)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

INVESTOR IN PEOPLE

OIPE JC161 JAN 22 2002 PATENT & TRADEMARK OFFICE

RECEIVED
FEB 04 2002
Technology Center 2600

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

CERTIFIED COPY OF PRIORITY DOCUMENT

Signed M Conkie

Dated 28 March 2001

1/77

Patents Form 1/77

The Patent Office

31MAY00 E540722-2 002073
P01/7700 0.00-0013011.2

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
B
26 MAY 2000
LONDON

The Patent Office

Cardiff Road
Newport
South Wales NP9 1RH

Statement of inventorship and of right to grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference — 9230

2. Patent application number *(The Patent Office will fill in this part)* — 26 MAY 2000 — 0013011.2

3. Full name, address and postcode of the or of each applicant (*underline all surnames*)

   NCR INTERNATIONAL, INC
   1700 SOUTH PATTERSON BOULEVARD
   DAYTON, OHIO 45479
   UNITED STATES OF AMERICA

   Patents ADP number (*if you know it*) — 07409352001

   If the applicant is a corporate body, give the country/state of its incorporation — INCORPORATED IN THE STATE OF DELAWARE

4. Title of the invention — METHOD AND APPARATUS FOR DETERMINING ONE OR MORE STATISTICAL ESTIMATORS OF CUSTOMER BEHAVIOUR

5. Name of your agent (*if you have one*)
   "Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

   F CLEARY
   INTERNATIONAL IP DEPARTMENT
   NCR LIMITED
   206 MARYLEBONE ROAD
   LONDON NW1 6LY

   Patents ADP number (*if you know it*) — 07704984001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (*if you know it*) the or each application number

Country	Priority application number (*if you know it*)	Date of Filing (*day/month/year*)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing (*day/month/year*)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (*Answer 'Yes' if:*
   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is not named as an applicant, or
   c) any named applicant is a corporate body.
   See note (d))

   YES

CERTIFIED COPY OF PRIORITY DOCUMENT

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document.

Continuation sheets of this form

Description	16
Claim(s)	3
Abstract	1
Drawing(s)	2 + 2

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translation of priority documents	
Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)	
Request for preliminary examination (*Patents Form 9/77*)	1
Request for substantive examination (*Patents Form 10/77*)	
Any other documents (*please specify*)	

11. I/We request the grant of a patent on the basis of this application.

Signature [signature] Date 26/05/2000

12. Name and daytime telephone number of person to contact in the United Kingdom

CHRISTINE SHEPPARD
020 7725 8379

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

DUPLICATE

METHOD AND APPARATUS FOR DETERMINING ONE OR MORE STATISTICAL ESTIMATORS OF CUSTOMER BEHAVIOUR

Background of the Invention

Field of the Invention

This invention relates to a method and apparatus for determining one or more statistical estimators of customer behaviour. The invention is particularly related to, but in no way limited to, modelling customer behaviour using a Bayesian statistical hidden Markov model technique.

Description of the prior art

Businesses typically have records of customer transaction histories. These records contain information that is potentially very valuable to the business because it enables the business to analyse customer behaviour and use this "feedback" to help plan the future of the business. However, assessments of the available data only provide information about customer behaviour that has already occurred. This is a drawback because behaviour patterns typically change over time. For example, a customer who is at present not very profitable could become more profitable in the future. There is thus a need to predict the future behaviour of customers.

One particular example concerns a business such as a bank which wishes to predict when a customer is likely to leave the bank. In that case such a prediction would be extremely advantageous because it allows the bank to take action such as to give incentives to the customer to prevent them from leaving.

Bayesian statistical techniques have been used to "learn" or make predictions on the basis of a historical data set. Bayes' theorem is a fundamental tool for a learning process that allows one to answer questions such as "How likely is my

hypothesis in view of these data?" For example, such a question could be "How likely is a particular future event to occur in view of these data?"

Bayes theorem is written as :

$$P(H/data) = \frac{P(data/H)P(H)}{P(data)}$$

Which can also be written as:

$$P(H/data) \propto P(data/H) \bullet P(H)$$

Because P(data) is unconditional and thus does not depend on H.

The probability of H given the data, P(H/data) is called the posterior probability of H. The unconditional probability of H, P(H) is called the prior probability of H and the probability of the data given H, P(data/H) is called the likelihood of H. By using knowledge and experience about past data an assessment of the prior probability can be made. New data is then collected and used to update the prior probability following Bayes theorem to produce a posterior probability. This posterior probability is then a prediction in the sense that it is a statement about the likelihood of a particular event occurring in the future. However, it is not simple to design and implement such Bayesian statistical methods in ways that are suited to particular practical applications.

It is accordingly an object of the present invention to provide a method and apparatus for determining one or more statistical estimators of customer behaviour, which overcomes or at least mitigates one or more of the problems noted above.

Summary of the Invention

According to an aspect of the present invention there is provided a method of determining one or more statistical estimators of future customer behaviour comprising the steps of:-

- accessing data about past customer behaviour;
- generating a Bayesian statistical model using the data about the past customer behaviour; and
- using the model to generate one or more statistical estimators of future customer behaviour.

A corresponding computer system is provided for determining one or more statistical estimators of future customer behaviour comprising:-

- an input arranged to access data about past customer behaviour;
- a processor arranged to generate a Bayesian statistical model using the data about the past customer behaviour; and
- wherein said processor is further arranged to use the model to generate one or more statistical estimators of future customer behaviour.

A corresponding computer program is provided for controlling a computer system such that one or more statistical estimators of future customer behaviour are determined said computer program being arranged to control the computer system such that:-

- data about past customer behaviour is accessed;
- a Bayesian statistical model is generated using the data about the past customer behaviour; and
- using the model, one or more statistical estimators of future customer behaviour are generated.

This provides the advantage that the statistical estimators of future customer behaviour are obtained and these may be used by a business, for example, to improve its performance. The data about past customer behaviour may comprise information about customer transactions such as cash machine withdrawal frequency. By using the method future customer transactions can then be predicted.

Preferably the method further comprises accessing information about customer attributes and wherein said model is generated using the information about customer attributes. This gives the advantage that the model is improved and found to enable good statistical estimators of future customer behaviour to be produced. The customer attributes could be the age, sex and salary of customers for example.

It is also preferred that the model comprises a representation of the customer behaviour in the form of a hidden Markov model with a random number of states. Moreover, it is preferred that the step of generating the model comprises clustering the past customer behaviour data into a plurality of states. It has unexpectedly been discovered that this type of statistical model is particularly effective for modelling customer behaviour data such as information about bank customers.

Advantageously, the behaviour of each customer over time is represented as a path through a plurality of the states and wherein these paths are unobserved and are considered random. This enables the evolution of customer behaviour over time to be modelled and in this way predictions about future customer behaviour can then be obtained from the model.

Preferably, each state is characterised by a random state parameter and preferably the model uses multi-variate customer data. That is a plurality of customer attributes such as age, sex and salary are used. This enables the model to be more effective for customer data and for particular applications such as predicting the future behaviour of bank customers.

Further benefits and advantages of the invention will become apparent from a consideration of the following detailed description given with reference to the accompanying drawings, which specify and show preferred embodiments of the invention.

Brief description of the drawings

Figure 1 is a flow diagram of a method of generating statistical estimators of customer behaviour.

Figure 2 is a flow diagram showing more detail about the step of generating a Bayesian statistical model from Figure 1.

Figure 3 is schematic diagram of a path between states which represents a customer's behaviour over time.

Figure 4 is a schematic diagram of a computer system.

Detailed description of the invention

Embodiments of the present invention are described below by way of example only. These examples represent the best ways of putting the invention into practice that are currently known to the Applicant although they are not the only ways in which this could be achieved.

Consider a business such as a bank. This bank may have beliefs, experience and past data about customer transactions. Using this information the bank can form an assessment of the prior probability that a particular customer will exhibit a certain behaviour, such as leave the bank. The bank may then collect new data about that customer's behaviour and using Bayes' theorem can update the prior probability using the new observed data to give a posterior probability that the customer will exhibit the particular behaviour such as leaving the bank. This posterior probability is a prediction in the sense that it is a statement of the likelihood of an event occurring. In this way the present invention uses Bayesian statistical

techniques to make predictions about customer behaviour. However, as mentioned above, it is not simple to design and implement such methods in ways that are suited to particular applications. The present invention involves such a method and is described in more detail below.

Figure 1 is a flow diagram of a method of determining statistical estimators of customer behaviour. Data about past customer behaviour is accessed (box 10 of Figure 1). For example, this data comprises information about customer transactions such as the frequency of cash withdrawals at a Bank's ATM machines and the amount of money withdrawn each time. Using this data a Bayesian statistical model is generated (see box 11 of Figure 1) and this model is then used to generate one or more statistical estimators of future customer behaviour (box 12 of Figure 1). As well as data about past customer behaviour, customer attributes such as age, sex and salary may be used to create the model.

The Bayesian statistical model that is used may be any suitable type of model which clusters the customer data and attributes into a finite number of states. Any suitable type of hidden Markov model technique may be used to achieve this.

In this way the Bayesian statistical model represents customer behaviour using a plurality of states (the number of which is unknown and considered random) where each state is characterised by a plurality of parameters. At a given point in time a customer's behaviour is represented using one of these states; that is the customer's behaviour at a particular time is a member of a particular state. All customers within a state are assumed to have behaviour that is homogeneous in some way. These states may be found to correspond to particular lifestyle groups such as employed single people, unemployed people, students etc. However, it may well also be the case that the clusters or states generated by the model do not correspond to lifestyle groups or other classes that are meaningful in social terms. In

order to represent a customer's behaviour over time, the model uses an unobserved path through these states. This is illustrated schematically in Figure 3. Time snapshots are represented by large circles 30 and within these clusters or states are represented by smaller black circles 31. Arrow 32 represents time. Suppose that a particular customer has behaviour at a first time that is represented by cluster 33 of the left most circle 30. The behaviour of that customer over time is then represented as a path between a state in each time shot circle 30. For example, Figure 3 shows such a path 33 for a customer who changes behaviour in each time shot. Thus customers are considered to move through different states over time, according to state transition probabilities, as their customer data and attributes evolve. In the statistical model used the paths of each customer through the states over time are not observed and are estimated or considered random. Also, each state k is characterised by a random state parameter $\theta^{k)}$. Observed customer transactions whilst they are in state k are assumed to follow a parametric probability model $p(Data|\,\theta^{k)})$.

A particular advantage of the present invention is that the model is arranged to deal with customer data comprising more than one parameter or attribute per customer. That is, the hidden Markov model technique used is arranged to use data that is not univariate. For example, a plurality of attributes for each customer (e.g. age, sex, salary) are used together with transaction data such as frequency of cash withdrawals from ATM machines. By using data that is multivariate (as opposed to univariate data) the model is improved such that the results are more accurate predictions of customer behaviour. As described below, Robert et al. (see section headed "references" below for full publication details) have described use of a hidden Markov model with a random number of states, but for only one time series of univariate data. Also, Robert et al did not consider applying these techniques to

customer data such as information about transactions and withdrawals from cash machines. It is not obvious that clustering techniques such as hidden Markov models are effective at dealing with such customer data and it has unexpectedly been discovered that the methods described herein are effective for such data.

Figure 2 is a flow diagram giving more detail about the step of generating the Bayesian statistical model. Bayesian prior probability distributions are specified for the number of states, the probabilities of a new customer starting in each state, the probabilities of moving between the different states and the state parameters (see boxes 21 to 23 of Figure 2). As already mentioned, the observed customer data is represented for each state using a parametric probability model (see box 24 of Figure 2). Using Bayes theorem, the Bayesian prior probability distributions, the accessed data and the parametric probability models are combined to generate a posterior probability distribution for each of:

- the number of states;
- the probabilities of a new customer starting in each state;
- the probabilities of moving between the different states; and
- the state parameters (see box 25 of Figure 2).

In the case that the unobserved state paths are treated as random, posterior probability distributions are also generated for these unobserved state paths.

The posterior probability distribution is then used to generate statistical estimators of future customer behaviour. For example, this may be done by using numerical or analytical methods to calculate the posterior probability distribution. Alternatively, and in a preferred embodiment, a sampling method is used to draw approximate random samples from the posterior distribution. Any suitable sampling method such as Gibbs sampling methods may be used. Once the samples have been drawn Monte Carlo inference is analysed using the samples to generate the

statistical estimators. For example, marginal distributions and predictive densities can be performed.

In the case that the customer data comprises information about transactions, the method gives outputs such as probabilities that particular customers will enter into certain transactions. For example, if the customer is a bank customer, the probability that a customer will leave a bank at a certain time can also be estimated. In this way an estimate of the lifetime value of that customer to the bank can be gained.

A detailed example of the method is now described:

Suppose there are R reference customers with whom the customer relationship has now ended and C current customers, and so $N = R + C$ customers overall. Then for each customer $i = 1,\ldots,N$, let n_i be the number of time units (e.g. weeks) over which transactions have been recorded. It is assumed that there are three observation types; a vector of attributes, W_i, that do not vary over time (e.g. the customer's sex); a matrix with n_i columns of attributes, X_i, which change over time but in a deterministic way (e.g. the customer's age each week); and a matrix with n_i columns of transactions, Y_i, which change over time in a non-deterministic way (e.g. the number of ATM visits made by a customer each week).

The evolution of customer behaviour is represented as a hidden Markov model (HMM) with a random number of states as described in Robert et al (2000). This model says that at any point in time a customer can be described as falling into one of a finite number of sets, and that within states customers will behave in some homogenous way. The number of states n is taken to be unknown and a Bayesian prior distribution is assigned. One choice would be n distributed uniformly between $\{2,3,\ldots,n_{\max}\}$. It is not essential to assume that the number of states is uniformly distributed in this way. Any other suitable distribution for the number of states may

be chosen. Each customer transaction history can then be viewed as dependent on an unobserved path z_l of length n_l through these states.

The Markov model is completed by the specification of an $n \times n$ transition probability matrix P with p_{ij} the probability of moving from state i to state j. State n is fixed to be the "end" state, representing the end of the customer relationship. Once entered this state cannot be left, so $p_{nn} = 1$ and $p_{nj} = 0$ for $j \neq n$. No transactions can be observed in this state.

One choice of prior distribution is to assume that, for $i = 1, \ldots, n-1$, the ith row p_i of the matrix P follows a Dirichlet distribution with parameter vector ξ_i. This provides the choice of setting $\xi_{ii} >> \xi_{ij}$ for $j \neq i$ to make remaining in one's present state much more likely than moving. Write π for the stationary distribution of P, so the probability of being at state i at a randomly selected time is π_i.

It is not essential to use a Dirichlet distribution as described above. Any other suitable distribution could be used. For example, a (n-1) variate normal distribution that is truncated so that each element lies between 0 and 1 and so that its sum is less than or equal to 1 could be used. By using a Dirichlet distribution computational advantages are achieved and it is simple to specify that a customer has a high probability of staying the same state between consecutive “time shots”.

If the records of a particular customer start at a random time into the customer relationship, the probability of that customer being in state i when the records commence is π_i.

If, on the other hand, the records start at the beginning of a customer relationship, then the initial state of the customer might have a different probability distribution, as some states may be more typical than otherwise for customers with

whom the relationship has just commenced. Write q_j for the probability of a new customer being in state j, $j = 1,\dots,n-1$. For a prior distribution, again one choice is to assume that the vector of probabilities $q = (q_1,\dots,q_{n-1})$ follows a Dirichlet distribution with parameter ξ_0.

For each customer i, define an identifier b_i which takes the value 1 if the records begin at the start of the customer relationship and 0 otherwise.

Now for each customer $i = 1,\dots N$, let $T_i = \{k | k \in \{1,\dots,n\}, \exists j \in \{1,\dots,n_i\} s.t. z_{ij} = k\}$ be the set of states visited by that customer, and let $S_{ki} = \{j | j \in \{1,\dots,n_i\}, z_{ij} = k\}$ be the (possibly empty) set of time indices j which customer i spends in state k. Note that $n \in T_i$ if and only if customer i is one of the R reference customers with whom the customer relationship has ended, and that $S_{ni} = \{n_i\}$ for reference customers and $S_{ni} = \varnothing$ otherwise.

Then for each state k define parameter vectors of length r $\theta^{(k)} = (\theta_1^{(k)},\dots,\theta_r^{(k)})$ to model the data via suitable parametric models. If conditional independence between customer observations given the parameters is assumed, and if a customers' transactions are also assumed conditionally independent given the parameters, the likelihood function is then given by

$$p(W,X,Y,z|n,P,q,\theta) = \prod_{i=1}^{N}\left\{q_{z_{i1}}^{b_i}\,\pi_{z_{i1}}^{(1-b_i)}\prod_{j=1}^{n_i-1}p_{z_{ij}z_{ij+1}}\,p(W_i,X_i|\,\theta^{k)},k\in T_i)\prod_{j=1}^{n_i}p(Y_{ij}|\,\theta^{z_{ij})})\right\}.$$

$$=\prod_{i=1}^{N}q_{z_{i1}}^{b_i}\,\pi_{z_{i1}}^{(1-b_i)}\prod_{k=}^{n-1}\prod_{l=1}^{n}p_{kl}^{m_{kl}}\prod_{i=1}^{N}\left\{p(W_i,X_i|\,\theta^{k)},k\in T_i)\prod_{j=1}^{n_i}p(Y_{ij}|\,\theta^{z_{ij})})\right\}$$

where $m_{kl} = \sum_{i=1}^{N} \sum_{j=1}^{n_i - 1} I\{z_{ij} = k, z_{ij+1} = l\}$ is the total number of times customers changed from state k to state l.

One choice of prior distribution of the $\theta^{k)}$ parameters which enables modelling of possible similarities between states through sharing common components, is to use a product of independent Dirichlet processes (see Ferguson, 1973; West et al, 1994). That is, for component $i = 1, \ldots r$,

$$\theta_i^{1)}, \ldots, \theta_i^{n)} \sim DP(\alpha_i F_i)$$

where α_i is a scalar precision parameter and F_i is a base prior which incorporates any prior beliefs that may be held about the distribution of the corresponding parameter component. However, it is also possible to use any other suitable prior distribution.

Bringing this all together, Bayes Theorem gives the posterior distribution of the parameters up to proportionality by

$$p(n, P, q, z, \theta | W, X, Y) \propto \prod_{i=1}^{N} q_{z_{i1}}^{b_i} \pi_{z_{i1}}^{(1-b_i)} \prod_{k=1}^{n-1} \left\{ q_k^{\alpha_{0k}-1} \prod_{l=1}^{n} p_{kl}^{\alpha_{kl}+m_{kl}-1} \right\} \prod_{i=1}^{N} \left\{ p\left(X_i, W_i | \theta^{k)}, k \in T_i\right) \prod_{j=1}^{n_i} p\left(Y_{ij} | \theta^{z_{ij})}\right) \right\}$$
$$\times \prod_{i=1}^{r} \prod_{k=1}^{n} \left\{ \left(\alpha_i + k - 1\right)^{-1} \alpha_i dF_i(\theta_i^{k)}) + \left(\alpha_i + k - 1\right)^{-1} \sum_{l=1}^{k} \delta(\theta_i^{l)}) \right\}$$

where $\delta(x)$ is a discrete probability mass function placing all its mass on x, and dF_i is the probability density/mass function of the distribution F_i. The constant of proportionality is the inverse of the multiple integral of the right hand side of the equation above with respect to $\{n, P, q, \theta, z\}$. Analytic calculations with the posterior

distribution are therefore complex. In a preferred embodiment, Markov Chain Monte Carlo (MCMC) simulation is used to draw approximate random samples from the posterior distribution for making parameter inference and prediction. However, this is not essential, any other suitable numerical method or analytic methods of calculating the posterior distribution may be used.

In a preferred embodiment, MCMC simulation is used as described above. For example, Gibbs sampling techniques are used. The Gibbs sampler is a MCMC technique for generating from the posterior distribution of a set of model parameters via the full conditional distributions. For a description of the Gibbs sampler and full conditional distributions see Smith and Roberts (1993). Two methods using Gibbs sampling are combined here.

The first was described by Robert et al (2000) for a HMM with a random number of states, but for only one time series of univariate data; the vector parameters $\left\{\theta^{(1)},\ldots,\theta^{(n)}\right\}$ are thus replaced by scalar parameters $\left\{\sigma^{(1)},\ldots,\sigma^{(n)}\right\}$. Because the number of states n is considered random, the MCMC Reversible jump methods of Green (1995) are required to explore the variable dimension parameter space. The jump moves described by Robert et al (2000) are used here to change the number of dimensions, with the only change that methods for deleting or adding a $\sigma^{(k)}$ parameter are here performed identically for each component of $\theta^{(k)}$ in turn. The Dirichlet process prior across states for corresponding components $\left\{\theta^{(1)},\ldots,\theta^{(n)}\right\}$ provides the advantage that two states that are to be merged have positive probability of already sharing common θ components and thus such a move will be

more likely to be accepted. The Gibbs moves for z and P (and here q) are identical to those described by Robert et al (2000).

To create a Gibbs move for the parameters $\left\{\theta^{1)},\ldots,\theta^{n)}\right\}$ conditional on $\{n,P,z\}$, the Gibbs sampling strategy of MacEachern (1992) for Dirichlet processes is implemented. However it is not essential to use this particular Gibbs sampling strategy. Any other suitable sampling methods can be used.

Once a large approximate sample from the posterior distribution $\{n,P,q,\theta,z\}^{(1)},\ldots,\{n,P,q,\theta,z\}^{(M)}$ has been collected, Monte Carlo inference about aspects of the posterior distribution such as marginal distributions and predictive densities can be performed. Thus predictions of customer transactions, how long the customer relationship will last and their lifetime value are all readily available.

The method described herein may be implemented using any suitable programming language executed on any suitable computing platform. For example, Matlab (trade mark) may be used together with a personal computer. A user interface is provided such as a graphical user interface to allow an operator to control the computer program, for example, to adjust the model, to display the results and to manage input of customer data. Any suitable form of user interface may be used as is known in the art.

Figure 4 is a schematic diagram of a computer system for generating statistical estimators of future customer behaviour. Data about past customer behaviour 42 is input to a processor 43 via an input 41. The processor uses this data to generate a Bayesian statistical model and using this model to generate statistical estimators 44 of future customer behaviour.

A range of applications are within the scope of the invention. These include situations in which it is required to determine one or more statistical estimators of customer behaviour. For example, to estimate the probability that a particular customer of a business will stop being a customer (for example by leaving a bank) at a specified time in the future or to estimate the frequency and nature of future customer transactions. Using such estimates the lifetime value of particular customers to a business can be estimated.

References

Ferguson, T. S. (1973) A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209-230.

Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian Model determination. *Biometrika* **82**, 711-732.

MacEachern, S. M. (1992) Estimating normal means with a conjugate style Dirichlet process prior. Technical report No 487, Department of Statistics, The Ohio State University.

Robert, C. P., Ryden, T., and Titterington, D. M. (2000) Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society Series B - Statistical Methodology* **62**, 57-75.

Smith, A. F. M. and Roberts, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B - Statistical Methodology* **55**, 3-23 (with discussion).

West, M., Mueller, P. and Escobar, M. D. (1994) Hierarchical priors and mixture models with applications in regression and density

estimation. *Aspects of Uncertainty: a Tribute to D. V. Lindley* (P. R. Freeman, and A. F. M. Smith, eds.). Chichester: Wiley.

Claims

1. A method of determining one or more statistical estimators of future customer behaviour comprising the steps of:-

(i) accessing data about past customer behaviour;

(ii) generating a Bayesian statistical model using the data about the past customer behaviour; and

(iii) using the model to generate one or more statistical estimators of future customer behaviour.

2. A method as claimed in claim 1 which further comprises accessing information about customer attributes and wherein said model is generated using the information about customer attributes.

3. A method as claimed in claim 1 or claim 2 wherein the model comprises a representation of the customer behaviour in the form of a hidden Markov model.

4. A method as claimed in claim 3 wherein said hidden Markov model has a random number of states.

5. A method as claimed in claim 1 wherein said step of generating the model comprises clustering the past customer behaviour data into a plurality of states.

6. A method as claimed in claim 5 wherein the behaviour of each customer over time is represented as a path through a plurality of the states and wherein these paths are unobserved and are considered random.

7. A method as claimed in any of claims 4 to 6 wherein each state is characterised by a plurality of random state parameters.

8. A method as claimed in claim 7 wherein past data about a customer's behaviour whilst that customer is in a particular state is assumed to follow a parametric probability model.

9. A method as claimed in any preceding claim wherein said step of generating the Bayesian statistical model comprises specifying a plurality of Bayesian prior probability distributions.

10. A method as claimed in claim 9 wherein said step of generating the model further comprises generating a plurality of Bayesian posterior probability distributions on the basis of at least the plurality of Bayesian prior probability distributions and the past customer data.

11. A method as claimed in claim 1 wherein said step (iii) of using the model to generate one or more statistical estimators comprises the step of using a sampling method to draw approximate random samples from the posterior distribution and performing Monte Carlo inference using the samples to generate the statistical estimators.

12. A method as claimed in claim 1 wherein said step (iii) of using the model to generate one or more statistical estimators comprises the step of numerically or analytically calculating the Bayesian posterior probability distributions.

13. A method as claimed in any preceding claim wherein the statistical estimators comprise a probability that a customer will exhibit a certain behaviour.

14. A method as claimed in any of claims 1 to 12 wherein the statistical estimators comprise the most probable behaviour exhibited by customers.

15. A method as claimed in any preceding claim wherein the past customer data comprises information about customer transactions.

16. A computer system for determining one or more statistical estimators of future customer behaviour comprising:-

(i) an input arranged to access data about past customer behaviour;

(ii) a processor arranged to generate a Bayesian statistical model using the data about the past customer behaviour; and

(iii) wherein said processor is further arranged to use the model to generate one or more statistical estimators of future customer behaviour.

17. A computer system as claimed in claim 16 wherein said data about past customer behaviour comprises customer attributes.

18. A computer system as claimed in claim 16 or claim 17 wherein the processor is arranged to generate the model by clustering the past customer behaviour data into a plurality of states.

19. A computer program for controlling a computer system such that one or more statistical estimators of future customer behaviour are determined said computer program being arranged to control the computer system such that:-

(i) data about past customer behaviour is accessed;

(ii) a Bayesian statistical model is generated using the data about the past customer behaviour; and

(iii) using the model, one or more statistical estimators of future customer behaviour are generated.

20. A computer program as claimed in claim 19 wherein said data about past customer behaviour comprises customer attributes.

21. A computer program as claimed in claim 19 or claim 20 which is arranged to control the computer system such that the processor generates the model by clustering the past customer behaviour data into a plurality of states.

ABSTRACT

Method and apparatus for determining one or more statistical estimators of customer behaviour

Businesses typically have large amounts of data about customer transactions and other customer information which is not fully utilised. The present invention provides a means of using this information to make predictions about future customer behaviour, for example by estimating the probability that a customer will leave a bank. Using these predictions the business is able to take action in order to improve its performance. Using customer data a Bayesian statistical model is generated and this model used to generate statistical estimators of customer behaviour. The statistical model is formed using hidden Markov model techniques by clustering customer data and attributes (e.g. Age, sex, salary) into a finite number of states. The number of states is unobserved and considered random. Bayesian prior probability distributions are specified and combined with the data to produce Bayesian posterior probability distributions. Using these Bayesian posterior probability distributions the statistical estimators are obtained. For example, Monte Carlo sampling techniques are used or alternatively the posterior distributions are calculated numerically or analytically.

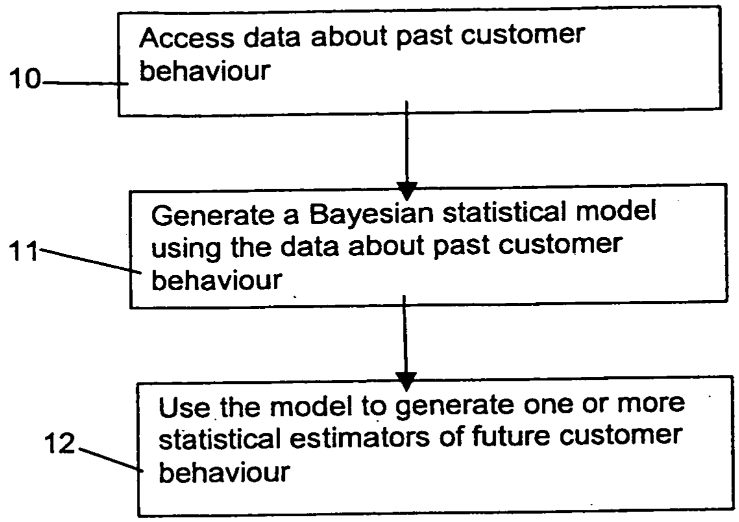

Figure 1

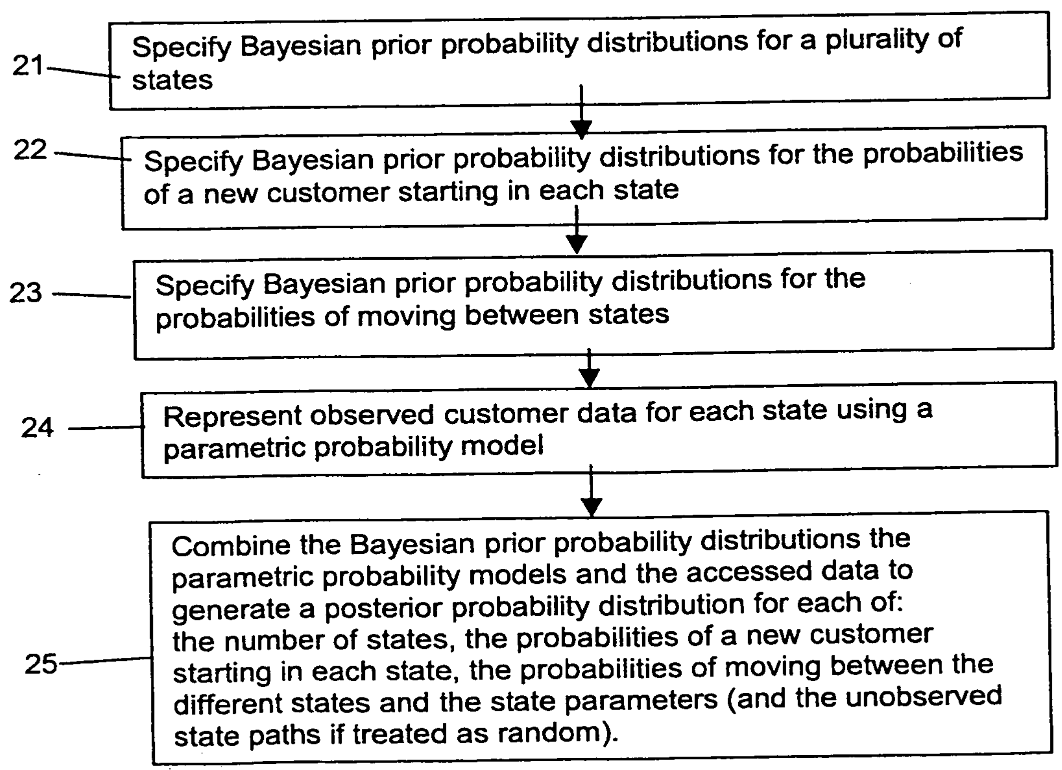

Figure 2

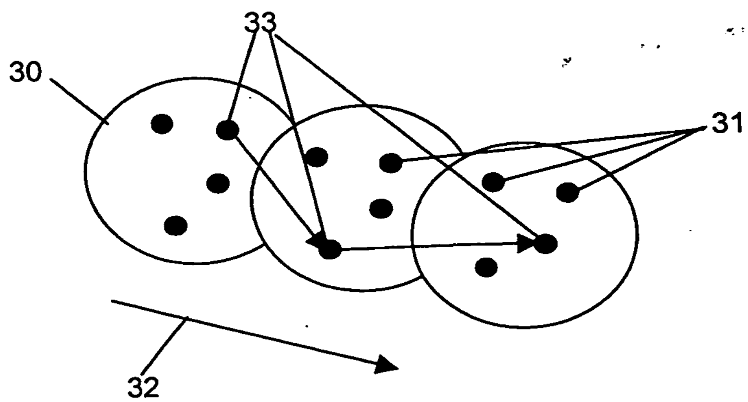

Figure 3

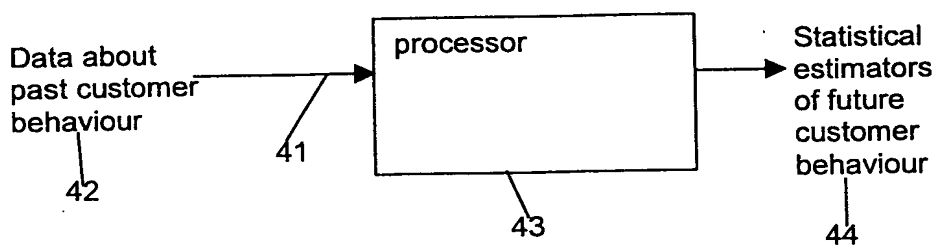

Figure 4